



On the Performance of Joint-Core Carrier-Phase Estimation in the Presence of Intercore Skew

Downloaded from: <https://research.chalmers.se>, 2023-05-05 10:17 UTC

Citation for the original published paper (version of record):

Alfredsson, A., Agrell, E., Karlsson, M. et al (2019). On the Performance of Joint-Core Carrier-Phase Estimation in the Presence of Intercore Skew. *Journal of Lightwave Technology*, 37(20): 5291-5298.
<http://dx.doi.org/10.1109/JLT.2019.2931987>

N.B. When citing this work, cite the original published paper.

On the Performance of Joint-Core Carrier-Phase Estimation in the Presence of Intercore Skew

Arni F. Alfredsson, *Student Member, IEEE*, Erik Agrell, *Fellow, IEEE*,
Magnus Karlsson, *Senior Member, IEEE, Fellow, OSA*, and Henk Wymeersch, *Member, IEEE*

Abstract—The sharing of lasers in space-division multiplexed multicore-fiber transmission yields correlated phase noise across the spatial channels. As a result, system performance can be improved through the use of joint-core carrier-phase estimation (CPE). However, the presence of intercore skew can reduce the potential of such schemes. This paper studies the effects of skew on pilot-aided joint-core CPE, where via simulations, it is shown that joint-core processing can be made to perform similarly to or better than separate-core processing for any amount of skew. It is shown that for a given signal-to-noise ratio (SNR) and skew, the performance of joint-core CPE relative to separate-core CPE is highly dependent on the ratio of the light-source linewidth to the local oscillator (LO) linewidth. In general, the performance of joint-core CPE in the presence of skew improves as the LO linewidth decreases compared to the light-source laser linewidth, assuming that the spatial channels are digitally realigned at the receiver.

Index Terms—Carrier-phase estimation, coherent communication, multicore fiber, skew, space-division multiplexing

I. INTRODUCTION

Phase noise is an inherent problem in coherent fiber-optic communication systems stemming from the imperfect nature of the lasers used as light sources at the transmitter and local oscillators (LOs) at the receiver. The time-varying nature of phase noise demands dynamic tracking, referred to here as carrier-phase estimation (CPE), which nowadays is typically carried out using digital signal processing (DSP) algorithms operating with the help of pilots [1], [2] or in a blind fashion without the use of pilots [3], [4].

CPE has in recent years been studied specifically in the context of multichannel fiber-optic systems [2], [5]–[8], such as in space-division multiplexed (SDM) transmission. SDM is believed to have the potential to upscale future optical networks in a cost-efficient way through, e.g., the integration of hardware and other resources, as well as the use of specific fibers such as multicore fibers (MCFs) and multimode fibers [9]. As a result, various transmission impairments become common to the spatial channels, which can be exploited. In particular, the sharing of lasers among channels will lead to correlated phase noise. This correlation can be used to

lower the required computational complexity of CPE through specialized transmission schemes [10] or DSP techniques such as master-slave CPE [8]. It also allows for the use of joint-channel CPE algorithms that outperform traditional single-channel CPE methods in terms of phase-noise tolerance.

Joint-channel processing for CPE has long been studied for wireless multiple-input multiple-output communications [11]–[13] and more recently for optical communications. In [14], an algorithm for joint-polarization CPE was proposed, where due to the high degree of correlation in the phase noise in both polarizations, significant performance gains were achieved over per-channel processing. Furthermore, in [2], [15], algorithms performing CPE jointly on an arbitrary number of channels were proposed and found effective based on experimental verification involving MCF transmission [5] and frequency-comb-based wavelength-division multiplexed (WDM) transmission [2].

In MCF systems, the optical signals do not propagate at identical velocities through the different cores due to variations in the refractive index between the cores. This results in relative delays between the received signals, a phenomenon typically called intercore skew [16]. Depending on the fiber design, typical skew values are on the order of 100 ps/km up to ns/km [17]. A similar observation is found in the case of transmission through a multimode fiber due to mode-dependent group delays, as well as in WDM transmission, where the received wavelength channels experience relative delays due to chromatic dispersion. Unless compensated optically before reaching the LO, these delays can reduce the phase-noise correlation across the channels, which in turn decreases the potency of schemes exploiting the correlation. This has been studied in the context of self-homodyne detection and master-slave CPE for MCF transmission [18], [19], as well as for frequency-comb WDM transmission [20]. In these studies, skew was found to be detrimental to the performance of the schemes. However, a similar analysis has yet to be carried out for joint-core CPE.

In [15], a simplistic multichannel phase-noise model was considered and used to develop a pilot-aided algorithm that performs joint-channel CPE for arbitrarily correlated phase noise. Both the model and the resulting algorithm proved effective when experimentally validated for transmission through a weakly-coupled, single-mode, homogeneous MCF [5]. The setup included careful path alignment to minimize any intercore skew, and hence, the effects of skew on the CPE performance were not included in the study.

As the CPE algorithm from [5] operates in a significantly

A. F. Alfredsson, E. Agrell, and H. Wymeersch are with the Department of Electrical Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden (e-mail: arnia@chalmers.se; agrell@chalmers.se; henkw@chalmers.se).

M. Karlsson is with the Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology SE-41296 Göteborg, Sweden (e-mail: magnus.karlsson@chalmers.se).

This work was supported by the Swedish Research Council (VR), Grants 2014-6138 and 2018-03701.

different manner than the phase-noise mitigation schemes of [18]–[20], the impact of skew on its performance cannot be directly inferred from [18]–[20]. Hence, in this paper, the following question is addressed: Can a joint-core CPE algorithm that was not specifically designed to account for intercore skew be adapted to become effective in the presence of skew? To answer this question, the phase-noise model and pilot-aided CPE algorithm from [5] are modified to account for intercore skew and used to evaluate skew-induced performance penalties. Preliminary results are presented in [21], where it is shown that by applying a small modification involving a numerical optimization to the CPE algorithm, joint-core processing can be made to perform similarly to or better than separate-core processing in the presence of skew. In particular, the results show that it is more beneficial for joint-core CPE to have a smaller LO linewidth compared to the light-source laser linewidth, assuming that the received signals are aligned in the receiver DSP. This paper expands on the results in [21] as follows: (i) A more comprehensive analysis of the system model is presented, illustrating why the considered CPE algorithm requires modification when skew is accounted for in the system model; (ii) Two methods are presented for modifying the algorithm, one entailing a numerical optimization and the other based on a heuristic argument. The two methods are compared in terms of performance and it is shown that in the presence of significant skew, they perform similarly; (iii) The performance of joint-core CPE is assessed in more detail in terms of the mean squared error (MSE) of the phase-noise estimates as a function of various system parameters, whose values for which joint-core CPE is beneficial are determined. The resulting bit error rate (BER) and achievable information rate (AIR) performance results for different number of cores are also presented.

II. SYSTEM MODEL

Consider uncoded transmission through D cores of an uncoupled-core, single-mode MCF, comprising a total of $2D$ parallel channels. All cores share the same light-source laser at the transmitter and the same LO at the receiver, and thus, the phase noise stemming from the lasers is common to all channels. The cores in the MCF have varying refractive indices, resulting in different propagation velocities that cause intercore skew between the received signals [16]. Interpolarization skew is considered negligible, which is a reasonable assumption [22] that matches experimental observations [5]. Effects due to the fiber Kerr nonlinearity are also neglected. In each channel, blocks of N symbols are transmitted. Each symbol is a point in $\mathcal{X} \in \mathbb{C}$, where \mathcal{X} is a set of constellation points corresponding to the modulation format, normalized such that the average symbol energy is E_s . Certain symbols within the transmitted blocks are designated as pilots whose positions and values are known to the receiver. Data symbols are modelled as uniform random variables that take on values from \mathcal{X} .

It is assumed that the received signal has undergone all standard DSP steps on a per-core basis, including dispersion compensation, phase-insensitive adaptive equalization [23],

the removal of frequency offsets¹, frame synchronization, and downsampling to one sample per symbol. The frame synchronization is assumed to include an ideal estimation of the intercore skew, and the signals are realigned in DSP on the receiver side. As a result, the phase noise from the light-source laser is also realigned across all channels but the LO phase noise becomes misaligned between channels on different cores. With these assumptions in place, the complex signal in polarization $w \in \{x, y\}$ of the optical field in core $i = 1, \dots, D$ is modelled in baseband and discrete time as

$$r_{w,k}^{(i)} = s_{w,k}^{(i)} e^{j\theta_{w,k}^{(i)}} + n_{w,k}^{(i)} \quad (1)$$

where $s_{w,k}^{(i)}$, $\theta_{w,k}^{(i)}$, and $n_{w,k}^{(i)}$ are the transmitted symbol, phase noise, and complex additive white Gaussian noise with variance $N_0/2$ per real dimension, resp., at time $k = 1, \dots, N$. It is also assumed that the channels are ordered as $(1, x), (1, y), (2, x), \dots, (D, y)$, where (i, w) denotes the channel corresponding to polarization w in core i .

In [5], the phase noise in each spatial channel was modelled as a sum of three components: a common laser phase noise, a core-specific phase drift, and a polarization-specific phase drift. All components were modelled as random walks. In this paper, the model is modified to account for intercore skew, which causes phase differences between cores that cannot be described as random walks.

First, all core- and polarization-specific phase drifts are assumed to be constants throughout each transmitted block. This is a reasonable assumption based on empirical evidence [5], [8]. The phase noise affecting the signal in polarization w of the optical field in core i at time k is then modelled as

$$\theta_{w,k}^{(i)} = \theta_{tx,k} + \theta_{rx,k+\xi_i}, \quad (2)$$

where ξ_i is the skew corresponding to core i with respect to an arbitrary time base and is known to the receiver from the frame synchronization. Note that since interpolarization skew is neglected, $\theta_{x,k}^{(i)} = \theta_{y,k}^{(i)}$ for all i, k . Furthermore, $[\theta_{tx,1}, \dots, \theta_{tx,N}]$ and $[\theta_{rx,1+\min_i \xi_i}, \dots, \theta_{rx,N+\max_i \xi_i}]$ are vectors of phase-noise samples from the light-source laser and LO, respectively. Both phase-noise processes are modelled as random walks,

$$\theta_{tx,k} = \theta_{tx,k-1} + \dot{\theta}_{tx,k} \quad (3)$$

$$\theta_{rx,k} = \theta_{rx,k-1} + \dot{\theta}_{rx,k}, \quad (4)$$

with $\theta_{tx,1}$ and $\theta_{rx,1}$ drawn uniformly on $[0, 2\pi)$, where the random-walk innovations, $[\dot{\theta}_{tx,2}, \dots, \dot{\theta}_{tx,N}]$ and $[\dot{\theta}_{rx,2+\min_i \xi_i}, \dots, \dot{\theta}_{rx,N+\max_i \xi_i}]$, are vectors of i.i.d. zero-mean Gaussian random variables with variances $2\pi\Delta\nu_{LS}T_s$ and $2\pi\Delta\nu_{LO}T_s$, with T_s denoting the symbol duration. Finally, $\Delta\nu_{LS}$ and $\Delta\nu_{LO}$ are the laser linewidths of the light-source laser and LO, respectively, and $\Delta\nu_C \triangleq \Delta\nu_{LS} + \Delta\nu_{LO}$ is the combined laser linewidth of the system.

As mentioned, the presence of intercore skew, which is compensated in DSP in the receiver, results in misalignment

¹When frequency-offset estimation is performed on a per-core basis, the compensated signals may exhibit residual relative frequency offsets between cores. These offsets were addressed in [5] and it was found that their impact can be made minimal, and hence, they are neglected in this paper.

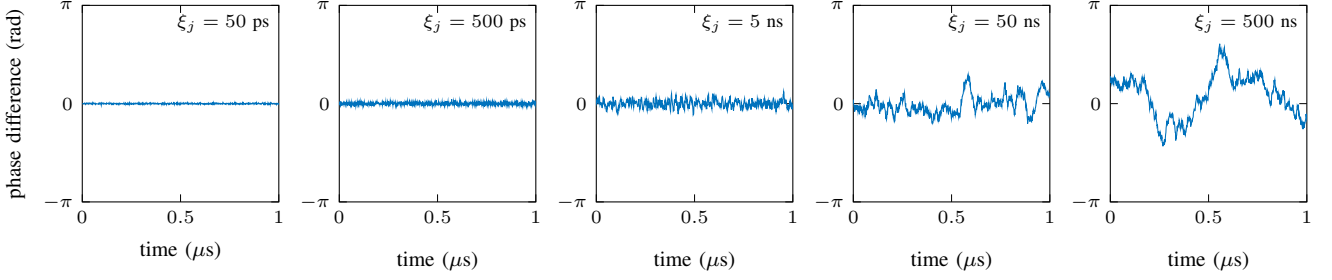


Fig. 1. The skew-induced phase difference between two arbitrary cores i and j when $\xi_i = 0$ and ξ_j takes on different values.

of the LO phase noise across the cores. The resulting phase difference between two arbitrary cores i and j is

$$\theta_{w,k}^{(i)} - \theta_{w,k}^{(j)} = \theta_{\text{rx},k+\xi_i} - \theta_{\text{rx},k+\xi_j} = b \sum_{l=k+\min(\xi_i, \xi_j)+1}^{k+\max(\xi_i, \xi_j)} \dot{\theta}_{\text{rx},l}, \quad (5)$$

which is a moving-average process [24, Ch. 10.4] of order $|\xi_i - \xi_j|$, with $b = 1$ if $\xi_i \geq \xi_j$ and $b = -1$ otherwise. Fig. 1 shows the skew-induced intercore phase differences (normalized to have zero mean) in (5) for 20 GBaud transmission with $\xi_i = 0$, ξ_j ranging from 50 ps (1 symbol) to 500 ns (10 000 symbols), and $\Delta\nu_{\text{LS}} = \Delta\nu_{\text{LO}} = 100$ kHz. As can be seen, the nature of the phase difference depends on the amount of intercore skew.

III. CPE STRATEGY

Two strategies are compared: (i) joint-core CPE and (ii) separate-core CPE. The former strategy is denoted by CPE_{JC} and entails CPE that is carried out jointly over channels in all cores and polarizations, whereas the latter strategy is denoted by CPE_{SC} and involves joint-polarization CPE that is performed independently in each core. The comparison between the two strategies is carried out in order to determine whether CPE_{JC} is beneficial compared to CPE_{SC} in the presence of skew. To that end, the model-based pilot-aided CPE algorithm from [5] is adapted to jointly process the channels in the presence of skew. It is originally designed based on the assumption that the signal has undergone the DSP steps that are listed in Sec. II. The algorithm iteratively computes the *a posteriori* probability distributions of the phase noise and data symbols, and uses the pilot symbols in the initial iteration to bootstrap the process. In this paper, a single iteration is run as it suffices to assess the effects of skew. Moreover, the algorithm operates on any number of channels based on the assumption that the phase noise across all channels is a multidimensional random walk, defined for $2D$ channels as

$$\begin{bmatrix} \phi_{1,k} \\ \vdots \\ \phi_{2D,k} \end{bmatrix} = \begin{bmatrix} \phi_{1,k-1} \\ \vdots \\ \phi_{2D,k-1} \end{bmatrix} + \begin{bmatrix} \dot{\phi}_{1,l} \\ \vdots \\ \dot{\phi}_{2D,l} \end{bmatrix} \quad (6)$$

where $[\phi_{1,1}, \dots, \phi_{2D,1}]^T$ is drawn uniformly on $[0, 2\pi)^{2D}$, and $[\dot{\phi}_{1,l}, \dots, \dot{\phi}_{2D,l}]^T$ is a zero-mean Gaussian random vector with a covariance matrix \mathbf{Q} that describes the phase-noise correlation across the channels. This covariance matrix is used by the algorithm when performing joint-channel CPE.

The phase difference between elements i and j of the left-hand side of (6),

$$\phi_{i,k} - \phi_{j,k} = \phi_{i,1} - \phi_{j,1} + \sum_{l=2}^k \dot{\phi}_{i,l} - \sum_{l=2}^k \dot{\phi}_{j,l}, \quad (7)$$

is in general a nonstationary process unless $\phi_{i,k} = \phi_{j,k}$ for all k since the variance of (7) increases with k . Moreover, $\theta_{w,k}^{(i)} - \theta_{w,k}^{(j)}$ in (5) is a stationary process by definition, and hence, the model in (2) cannot be described as a multidimensional random walk across the channels. However, due to the random-walk assumption, the CPE algorithm requires \mathbf{Q} to be estimated. It is found that the best covariance matrix depends on various parameters, such as the pilot rate, the distribution of the pilots in the transmitted symbol blocks, the intercore skew, and the ratio $R_{\text{LW}} \triangleq \Delta\nu_{\text{LS}}/\Delta\nu_{\text{LO}}$. Hence, the algorithm is adapted to use a particular matrix that is obtained using either a numerical optimization for each parameter setting such that the resulting MSE of the estimated phase noise is minimized, or a heuristic construction based on the system model without the need for any optimization. The covariance matrices resulting from the numerical optimization and the heuristic construction are denoted by \mathbf{Q}_{O} and \mathbf{Q}_{H} , respectively.

To facilitate the optimization, \mathbf{Q}_{O} is constrained to a particular structure based on certain properties of the system model, namely that the phase noise is perfectly correlated across the two polarizations in each core, but can be arbitrarily correlated across cores. For 2 cores, the covariance matrix is on the form

$$\mathbf{Q}_{\text{O}} = 2\pi\Delta\nu_{\text{C}}T_s \begin{bmatrix} \mathbf{1}_2 & \alpha\mathbf{1}_2 \\ \alpha\mathbf{1}_2 & \mathbf{1}_2 \end{bmatrix}, \quad (8)$$

where $\mathbf{1}_2$ is a 2×2 matrix of ones and $\alpha \in [0, 1]$ is an optimization parameter, with $\alpha = 0$ and $\alpha = 1$ corresponding to no and perfect correlation, respectively. The structure of \mathbf{Q}_{O} can be extended to D cores as

$$\mathbf{Q}_{\text{O}} = 2\pi\Delta\nu_{\text{C}}T_s \begin{bmatrix} \alpha_{1,1}\mathbf{1}_2 & \alpha_{1,2}\mathbf{1}_2 & \cdots & \alpha_{1,D}\mathbf{1}_2 \\ \alpha_{1,2}\mathbf{1}_2 & \alpha_{2,2}\mathbf{1}_2 & \cdots & \alpha_{2,D}\mathbf{1}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,D}\mathbf{1}_2 & \alpha_{2,D}\mathbf{1}_2 & \cdots & \alpha_{D,D}\mathbf{1}_2 \end{bmatrix}, \quad (9)$$

comprising an optimization parameter for each pair of cores, constrained such that \mathbf{Q}_{O} is positive semidefinite and $\alpha_{i,i} = 1$ for all i . Finally, CPE_{SC} , whose performance is independent of skew, is implemented by fixing $\alpha_{i,j} = 0$ for all $i \neq j$.

Our suggestion for a heuristic construction for any number of cores is computing the covariance matrix of $\theta_k - \theta_{k-1}$, where $\theta_k \triangleq [\theta_{x,k}^{(1)}, \theta_{y,k}^{(1)}, \theta_{x,k}^{(2)}, \dots, \theta_{y,k}^{(D)}]^T$, denoted as

$$\mathbf{Q}_H = 2\pi\Delta\nu_C T_s \begin{bmatrix} q_{1,1}\mathbf{1}_2 & q_{1,2}\mathbf{1}_2 & \cdots & q_{1,D}\mathbf{1}_2 \\ q_{1,2}\mathbf{1}_2 & q_{2,2}\mathbf{1}_2 & \cdots & q_{2,D}\mathbf{1}_2 \\ \vdots & \vdots & \ddots & \vdots \\ q_{1,D}\mathbf{1}_2 & q_{2,D}\mathbf{1}_2 & \cdots & q_{D,D}\mathbf{1}_2 \end{bmatrix}, \quad (10)$$

where it is shown in Appendix A that

$$q_{i,j} = \begin{cases} 1, & \xi_i = \xi_j \\ R_{LW}/(1 + R_{LW}), & \xi_i \neq \xi_j \end{cases}. \quad (11)$$

Thus, \mathbf{Q}_H can be regarded as a special case of \mathbf{Q}_O . Furthermore, it will be shown in Sec. IV that when the intercore skew grows large, the optimization in (9) yields approximately the same values as (10). Thus, even though \mathbf{Q}_H does in general not guarantee optimality² for the phase-noise model in (2), it is near optimal in the presence of significant skew.

It is worth noting that a potentially more effective joint-core CPE algorithm could be developed based on (2), which would take intercore skew into account by design. Moreover, it is not obvious how beneficial joint adaptive equalization and CPE would be compared with separately performing these tasks in the presence of skew. While out of scope for this paper, such algorithm designs would be interesting as future work for performance comparison.

IV. NUMERICAL RESULTS

In this section, skew-induced performance effects are assessed for uncoded transmission of quadrature amplitude modulation (QAM) at 20 GBaud through an uncoupled MCF via Monte Carlo simulations. To that end, the MSE of the estimated phase noise is measured for different light-source laser and LO linewidths, amounts of intercore skew, and signal-to-noise ratios (SNRs), denoted by E_s/N_0 . All the results pertain to 2-core transmission except in Fig. 10, where BERs and AIRs are presented for transmission through 2 and 10 cores. Each AIR is based on computing the generalized mutual information for the mismatched decoder developed in [15], and subtracting the pilot rate. Unless specified otherwise, $\Delta\nu_C = 200$ kHz, and for the results pertaining to 2 cores, $\xi_1 = 0$ and ξ_2 takes on different values; thus, the intercore skew is encapsulated in $\xi \triangleq |\xi_2 - \xi_1| = \xi_2$. A frame of at least 10 000 symbols is transmitted per channel, including both pilot and data symbols, and each result is obtained through repeated transmission until at least 10^7 symbols have been transmitted across all channels.

During the initial iteration of the CPE algorithm, the phase noise is estimated at the pilot positions and interpolated over time and channels using Kalman smoothing recursions. Hence, the accuracy of the resulting phase-noise estimates at the data-symbol positions is highly dependent on the distribution of the pilot symbols. The optimal distribution of pilot symbols is nontrivial to find in general, and it depends on various

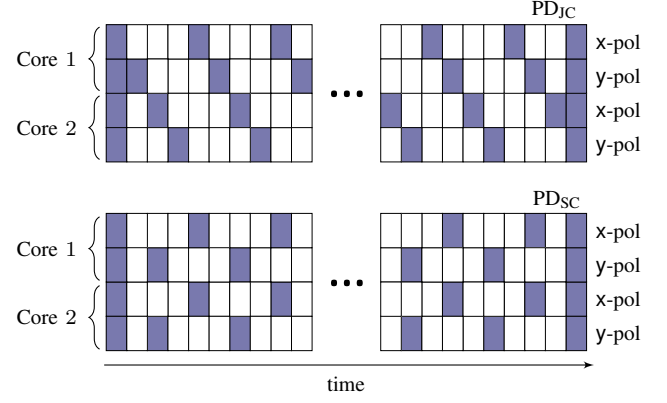


Fig. 2. The distributions of pilot symbols used by CPE_{JC} and CPE_{SC} with approximately 25% pilot rate for illustration purposes. Dark and white squares indicate pilot and data symbols, respectively.

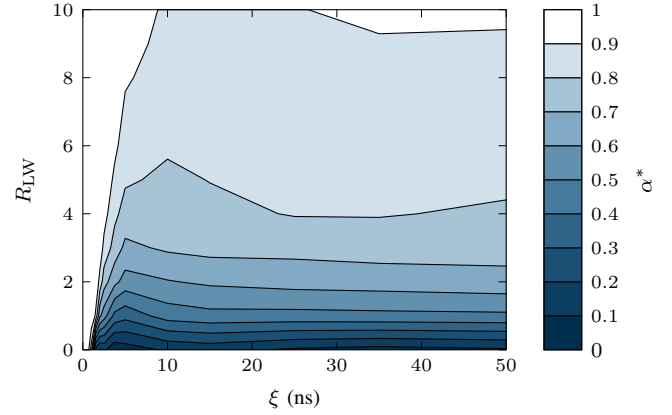


Fig. 3. A contour plot of the obtained α^* as a function of ξ and R_{LW} when $E_s/N_0 \rightarrow \infty$.

parameters such as E_s/N_0 , the number of channels, and the spatial correlation in the phase noise. Without claims of optimality, a heuristically constructed pilot distribution that was shown to be effective for CPE_{JC} in [25] is used unless stated otherwise, denoted by PD_{JC} and depicted in Fig. 2. In the case of CPE_{SC}, a pilot distribution denoted by PD_{SC} and also shown in Fig. 2 is used unless mentioned otherwise as it is found to be more effective for CPE_{SC} than PD_{JC}. For both pilot distributions, the pilot rate is controlled by the temporal spacing between consecutive pilots. In this paper, the pilot rate is set to 1%.

As mentioned in Sec. III, CPE_{JC} is realized by using either \mathbf{Q}_O or \mathbf{Q}_H . In the former case, the optimal value of α in (8) is computed for each system-parameter setting as $\alpha^* = \arg\min_{\alpha \in [0,1]} \text{MSE}(\alpha)$, implemented using 15 iterations of golden-section search. An example of the resulting α^* is depicted in Fig. 3 as a function of ξ and R_{LW} when $E_s/N_0 \rightarrow \infty$. Similar values of α^* are acquired overall for other tested E_s/N_0 . When $\xi = 0$, $\alpha^* = 1$, and as ξ grows large, $\alpha^* \approx R_{LW}/(1 + R_{LW})$ for all tested pilot rates. Therefore, when R_{LW} is low, skew severely decorrelates the phase noise across the cores and the best value of α is close to 0, which results in CPE_{JC} operating similarly to CPE_{SC}.

²If the phase noise is modelled as the multidimensional random walk in (6), the construction in (10) is optimal.

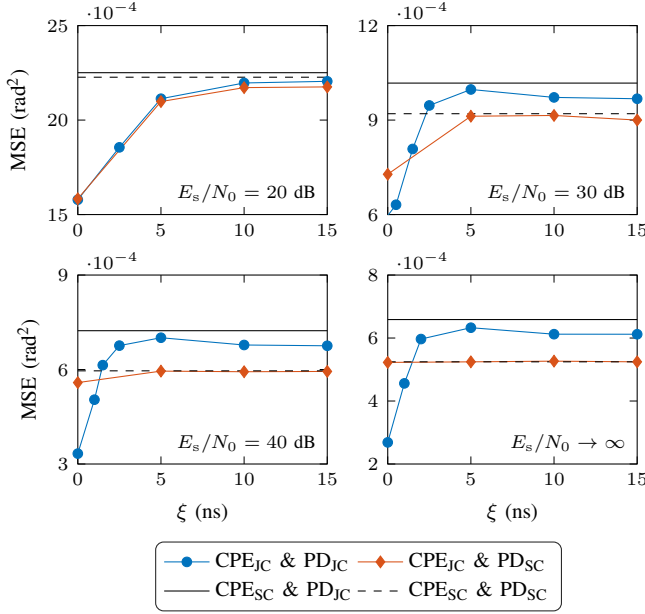


Fig. 4. Performance comparison of CPE_{JC} using \mathbf{Q}_O and CPE_{SC} for $R_{\text{LW}} = 1$ and different E_s/N_0 , illustrating the fact that given a particular pilot distribution, CPE_{JC} performs similar to or better than CPE_{SC} .

Moreover, it is found that the transition between $\alpha^* = 1$ and $\alpha^* \approx R_{\text{LW}}/(1 + R_{\text{LW}})$ accelerates as the pilot rate increases. Thus, \mathbf{Q}_O and \mathbf{Q}_H become more similar with increasing pilot rate. The performance results in Figs. 4–7 pertain to the performance of CPE_{JC} using \mathbf{Q}_O , whereas the results in Fig. 7–11 and Table I are obtained using \mathbf{Q}_H .

It is found that CPE_{JC} performs similarly to or better than CPE_{SC} when both strategies use the same pilot distribution. To demonstrate this, Fig. 4 shows the performance of CPE_{JC} and CPE_{SC} using both of the considered pilot distributions for $R_{\text{LW}} = 1$ and different E_s/N_0 . As can be seen, CPE_{JC} using PD_{SC} never sees worse performance than CPE_{SC} and outperforms CPE_{SC} at low values of E_s/N_0 and ξ . At $E_s/N_0 = 20$ dB, the performance of CPE_{JC} does not depend significantly on whether PD_{JC} or PD_{SC} is used. However, for negligible intercore skew, CPE_{JC} using PD_{JC} yields the lowest MSE. Similar results are found for all other tested R_{LW} .

Fig. 5 shows the MSE performance of CPE_{JC} as a function of ξ for different values of R_{LW} when $E_s/N_0 \rightarrow \infty$. In the absence of skew ($\xi = 0$), the performance does not depend on R_{LW} . As the amount of skew grows, however, the MSE sees an increase whose speed and amount depends on R_{LW} . In general, the performance improves with increasing R_{LW} , i.e., the benefits of CPE_{JC} grow with decreasing linewidth of the LO compared to the light-source laser. Included in the plot is the performance of CPE_{SC} , which is independent of skew, shown with a dashed line. The reason CPE_{SC} outperforms CPE_{JC} beyond a particular skew for low R_{LW} is the fact that it uses PD_{SC} , which is well suited for joint-polarization processing on a per-core basis. For large enough skew and low R_{LW} , α^* takes on such a small value that CPE_{JC} is essentially operating in the same manner as CPE_{SC} , but with a more suboptimal pilot distribution in PD_{JC} . Naturally, the

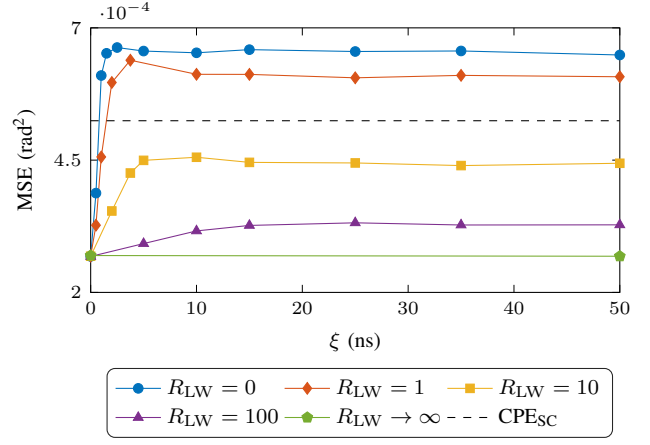


Fig. 5. The phase-noise estimate MSE performance of CPE_{JC} using \mathbf{Q}_O versus ξ when $E_s/N_0 \rightarrow \infty$, for different values of R_{LW} .

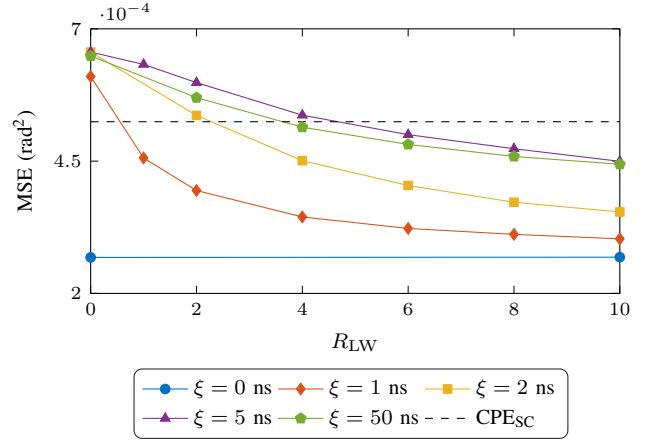


Fig. 6. The phase-noise estimate MSE performance of CPE_{JC} using \mathbf{Q}_O versus R_{LW} when $E_s/N_0 \rightarrow \infty$, for different values of ξ .

performance of CPE_{JC} could in principle be optimized over the two pilot distributions. The implementation of this scheme, however, would require feedback to the transmitter as the best pilot distribution depends on ξ among other parameters.

It is clear from Fig. 5 that CPE_{JC} is beneficial as long as R_{LW} is sufficiently large. Fig. 6 shows in more detail the MSE performance of CPE_{JC} as a function of R_{LW} for different values of ξ when $E_s/N_0 \rightarrow \infty$. As before, the performance of CPE_{SC} is shown with a dashed line for comparison. As R_{LW} increases, the performance of CPE_{JC} eventually exceeds the performance of CPE_{SC} . The point at which this occurs depends on ξ , with $\xi = 5$ ns representing the worst case out of the tested ξ , requiring $R_{\text{LW}} \approx 5$. Hence, for $R_{\text{LW}} > 5$, CPE_{JC} outperforms CPE_{SC} for any ξ . It is found that for $E_s/N_0 \in [20, 40]$ dB, a lower minimum value of R_{LW} is required for CPE_{JC} to outperform CPE_{SC} .

The curves in Fig. 7 show for different E_s/N_0 where the MSE of CPE_{JC} , when computed as a function of ξ and R_{LW} , equals the MSE of CPE_{SC} . The performance of CPE_{JC} using \mathbf{Q}_H is also shown with dashed curves as a comparison. As can be seen, the performance of CPE_{JC} does not depend significantly on whether \mathbf{Q}_H or \mathbf{Q}_O is used. The regions above

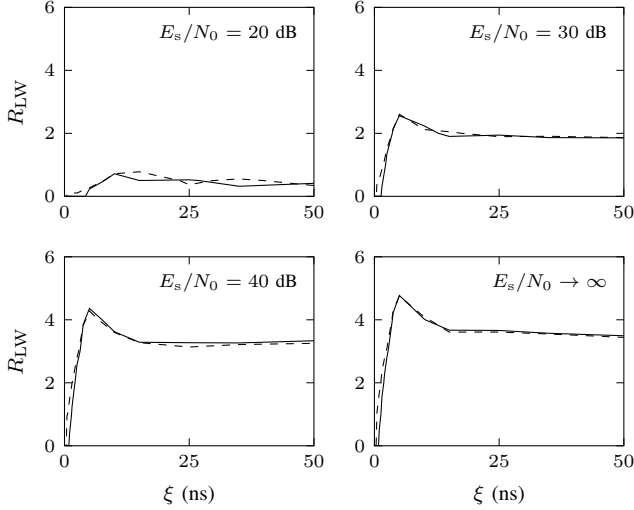


Fig. 7. Equal-performance curves of CPE_{JC} and CPE_{SC} as functions of ξ and R_{LW} for different E_s/N_0 , where solid and dashed curves represent CPE_{JC} using \mathbf{Q}_O and \mathbf{Q}_H , respectively. The areas above the curves correspond to values of ξ and R_{LW} where CPE_{JC} outperforms CPE_{SC} .

the curves correspond to values of ξ and R_{LW} where CPE_{JC} outperforms CPE_{SC} . At low E_s/N_0 , CPE_{JC} is beneficial for a larger range of R_{LW} , outperforming CPE_{SC} for $R_{\text{LW}} \geq 1$ at $E_s/N_0 = 20$ dB. However, as the E_s/N_0 grows large, CPE_{JC} is not beneficial for all ξ unless $R_{\text{LW}} \geq 5$. Moreover, it is found that the pilot rate does not influence the results significantly; CPE_{JC} becomes more sensitive to an increase in ξ but is beneficial for a slightly larger range of R_{LW} as the pilot rate grows.

Fig. 8 further compares the MSE performance of CPE_{JC} using \mathbf{Q}_H with the MSE performance of CPE_{SC} . To visualize the difference, the relative MSE performance metric $\max_{\xi} \text{MSE}_{\text{JC}}(\xi)/\text{MSE}_{\text{SC}}$ is plotted as a function of $\Delta\nu_{\text{LS}}$ and $\Delta\nu_{\text{LO}}$, where $\text{MSE}_{\text{JC}}(\xi)$ denotes the MSE of CPE_{JC} using \mathbf{Q}_H when computed as a function of $\xi \in [0, 50]$ ns, and MSE_{SC} denotes the MSE of CPE_{SC} . When $E_s/N_0 \rightarrow \infty$, the relative performance between CPE_{JC} and CPE_{SC} is invariant to $\Delta\nu_{\text{C}}$, which is demonstrated by the diagonal contour lines in the right plot of Fig. 8. Furthermore, at $E_s/N_0 = 25$ dB, the relative performance is approximately invariant to $\Delta\nu_{\text{C}}$ for high values of R_{LW} .

To assess the sensitivity to linewidth estimation errors, Fig. 9 shows the MSE performance of CPE_{JC} using \mathbf{Q}_H and CPE_{SC} versus the estimation error of $\Delta\nu_{\text{LO}}$ where $\Delta\nu_{\text{LS}}$ is assumed to be perfectly estimated, for $R_{\text{LW}} = 4$ and $E_s/N_0 \rightarrow \infty$. As evident, CPE_{JC} shows a greater sensitivity to estimation errors of $\Delta\nu_{\text{LO}}$ compared with CPE_{SC} . This is owing to the estimation error of $\Delta\nu_{\text{LO}}$ yielding inaccuracies in the values of $\Delta\nu_{\text{C}}$ and R_{LW} . As CPE_{JC} makes use of both of these parameters, it sees a larger penalty than CPE_{SC} , which uses only $\Delta\nu_{\text{C}}$.

It is of practical interest to determine how the MSE results presented so far translate to BER and AIR performance. Thus, Fig. 10 shows BER and AIR versus E_s/N_0 for transmission of 256QAM. Every core has a uniformly distributed random skew in the range $[0, 50]$ ns, and as each transmission sees

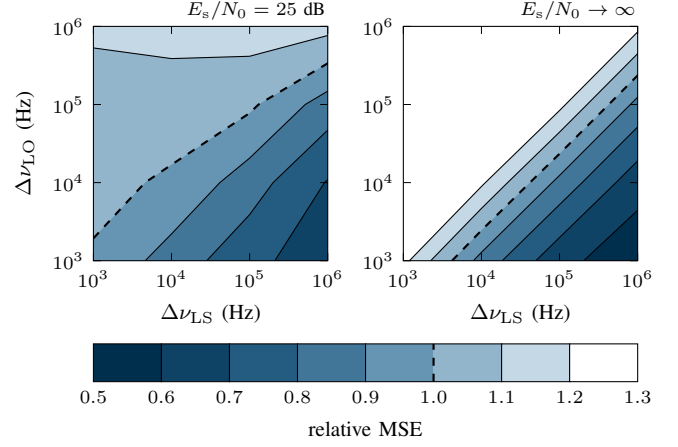


Fig. 8. The MSE performance of CPE_{JC} using \mathbf{Q}_H , relative to that of CPE_{SC} , versus $\Delta\nu_{\text{LS}}$ and $\Delta\nu_{\text{LO}}$ for $\xi = 50$ ns and different values of E_s/N_0 . Diagonal contour lines indicate that the relative MSE depends on R_{LW} but not on $\Delta\nu_{\text{C}}$.

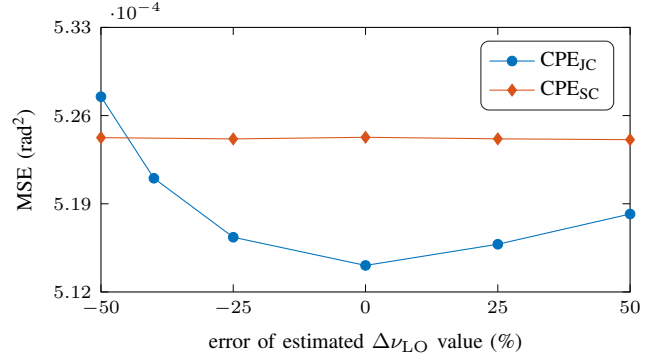


Fig. 9. The phase-noise estimate MSE of CPE_{JC} using \mathbf{Q}_H and CPE_{SC} versus the error of the estimated $\Delta\nu_{\text{LO}}$ value when the true value is 40 kHz, where $R_{\text{LW}} = 4$, $E_s/N_0 \rightarrow \infty$, and perfect estimation of $\Delta\nu_{\text{LS}}$ is assumed.

different realizations of the random skews, the performance results are averaged over at least 100 realizations. Moreover, the performance of CPE_{JC} depends on the number of cores that are processed, and thus, the results entail transmission through 2 and 10 cores. The performance of transmission in the absence of phase noise is included as a reference. To maintain a reasonable simulation time for both numbers of cores, CPE_{JC} is realized using \mathbf{Q}_H . From left to right, the rows of plots pertain to $R_{\text{LW}} = 0$, $R_{\text{LW}} = 1$, $R_{\text{LW}} = 10$, and $R_{\text{LW}} \rightarrow \infty$, with the corresponding values of $\Delta\nu_{\text{LS}}$ and $\Delta\nu_{\text{LO}}$ noted in the plots. The BER and AIR results are consistent with each other and are in agreement with Fig. 7. For $R_{\text{LW}} = 0$, CPE_{SC} sees better performance than CPE_{JC} , and for $R_{\text{LW}} = 1$, the two strategies perform similarly. As R_{LW} increases, however, the performance of CPE_{JC} improves with respect to CPE_{SC} , whose performance is independent of ξ and R_{LW} . Table I shows the increase in power efficiency (PE), defined here as the reduction in the required E_s/N_0 to attain a given AIR, and the increase in AIR per polarization, maximized over the studied E_s/N_0 range for different combinations of R_{LW} and number of cores. The biggest performance improvement is seen for 10 cores when $R_{\text{LW}} \rightarrow \infty$, i.e., an increase of 1.73 dB in PE or an

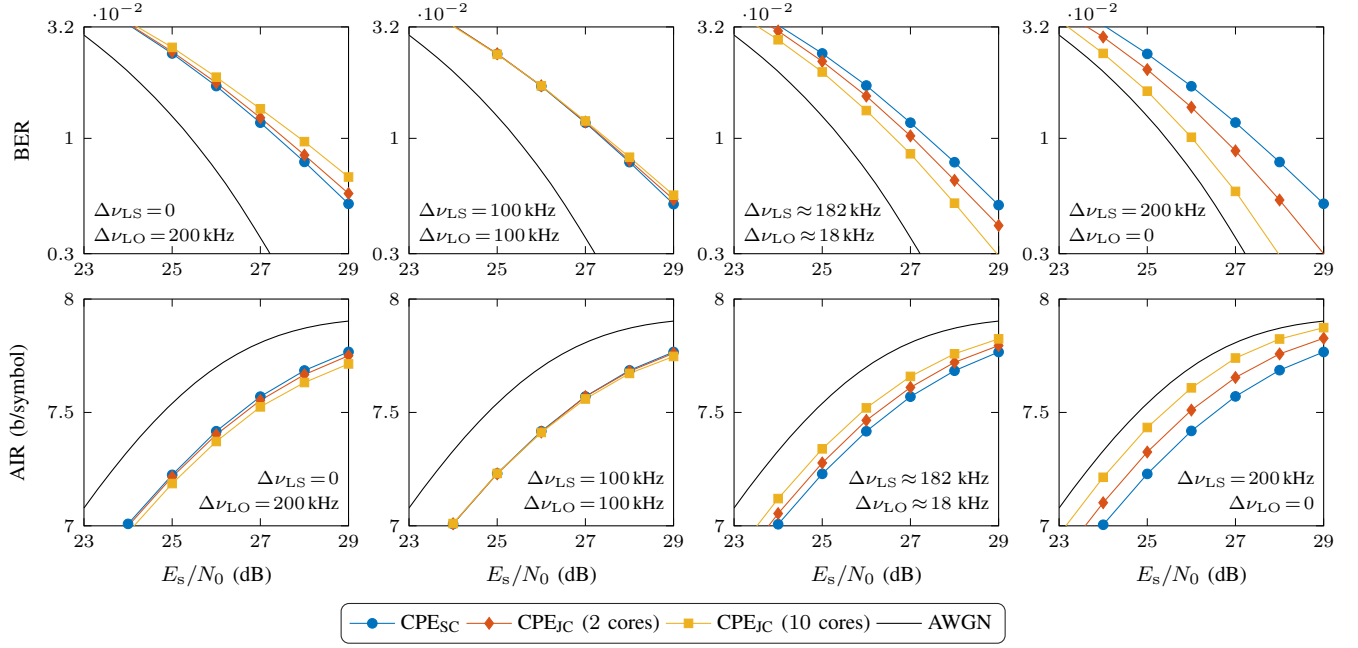


Fig. 10. BER and AIR versus E_s/N_0 for 256QAM transmission, comparing the performance of CPE_{JC} using \mathbf{Q}_H and CPE_{SC} for different R_{LW} and numbers of cores, where each core has a random skew in the range [0, 50] ns.

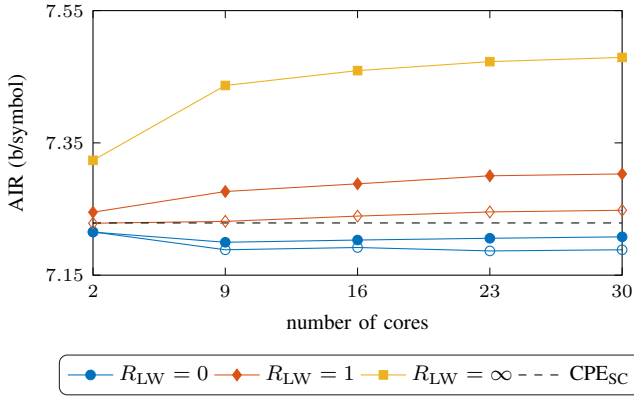


Fig. 11. The AIR performance of CPE_{JC} using \mathbf{Q}_H as a function of the number of cores for transmission of 256QAM at $E_s/N_0 = 25$ dB. Filled and empty curve markers correspond to each core having random skew in the range [0, 5] ns and [0, 500] ns, respectively.

TABLE I

THE MAXIMUM INCREASE IN PE AND AIR PER POLARIZATION THAT JOINT-CORE CPE ENABLES IN THE PRESENCE OF SIGNIFICANT SKEW

	PE increase (dB)		AIR increase (b/symbol)	
	2 cores	10 cores	2 cores	10 cores
$R_{LW} = 10$	0.45	0.90	0.05	0.11
$R_{LW} \rightarrow \infty$	0.90	1.73	0.10	0.21

increase of 0.21 b/symbol in AIR per polarization.

To study in greater detail how the performance of CPE_{JC} depends on the number of cores, Fig. 11 shows the AIR performance of CPE_{JC} using \mathbf{Q}_H versus the number of cores for transmission of 256QAM at $E_s/N_0 = 25$ dB. The perfor-

mance of CPE_{SC} is also included as a reference. For $R_{LW} \geq 1$, the performance of CPE_{JC} improves with an increasing number of cores for both considered skew conditions, whereas the opposite is observed for $R_{LW} = 0$.

V. DISCUSSION AND CONCLUSIONS

As phase noise is generally highly correlated across polarization-multiplexed channels, joint-polarization processing for CPE is usually beneficial compared to per-polarization processing. In MCF transmission, it has been previously shown that performing CPE jointly across all polarizations and cores can provide further performance improvements, provided that intercore skew is negligible. However, preventing skew may prove difficult in practice, and in the presence of skew on the order of ns in the received signal, the benefits of joint-core CPE can quickly diminish.

In this paper, the effects of skew on the performance of joint-core CPE were assessed in a quantitative manner and compared to the performance of separate-core CPE. To achieve this, a multichannel phase-noise model neglecting skew that was introduced in [5] and a pilot-aided joint-channel CPE algorithm based on this model were adapted to manage intercore skew. The adapted model and algorithm were then used to investigate the effect of skew on joint-core CPE. To summarize the findings, it was shown that joint-core CPE performs similarly to or better than separate-core CPE, regardless of skew (see Fig. 4). Moreover, for high SNRs and significant skew, the relative performance between joint-core and separate-core CPE depends on the ratio between the two laser linewidths but is invariant to the combined linewidth (see Fig. 8).

Assuming that the received signals have been realigned in the receiver DSP, joint-core CPE is beneficial for any skew

as long as the LO linewidth is narrower than the light-source laser linewidth by at least a factor of 1 to 5, depending on the SNR (see Figs. 5–7). The most significant performance improvements due to joint-core CPE, compared with separate-core CPE, were found for an ideal zero-linewidth LO. In this case, a PE increase of 0.9 dB and 1.7 dB, or alternatively, an increase of 0.1 b/symbol and 0.2 b/symbol in AIR per polarization was observed for 2-core and 10-core transmission, respectively (see Figs. 10–11 and Table I). However, joint-core CPE does not provide significant benefits in the presence of significant skew if the light-source laser linewidth is greater than the LO linewidth. Joint-core CPE is also found to be more sensitive to linewidth estimation errors compared with separate-core CPE (see Fig. 9). Hence, if separate-core CPE can be implemented with less computational complexity than joint-core CPE, it may be a more reasonable strategy in some cases when the intercore skew is not mitigated optically.

APPENDIX A DERIVATION OF HEURISTIC CONSTRUCTION

Let $E[\cdot]$ denote the expectation of a random variable. Using (2) and the definition of covariance, each $q_{i,j}$ of \mathbf{Q}_H in (10) is

$$q_{i,j} = \frac{E[(\theta_{w,k}^{(i)} - \theta_{w,k-1}^{(i)})(\theta_{w,k}^{(j)} - \theta_{w,k-1}^{(j)})]}{2\pi\Delta\nu_C T_s} \quad (12)$$

$$= \frac{E[(\dot{\theta}_{\text{tx},k} + \dot{\theta}_{\text{rx},k+\xi_i})(\dot{\theta}_{\text{tx},k} + \dot{\theta}_{\text{rx},k+\xi_j})]}{2\pi\Delta\nu_C T_s} \quad (13)$$

$$= \frac{E[(\dot{\theta}_{\text{tx},k})^2] + E[\dot{\theta}_{\text{rx},k+\xi_i}\dot{\theta}_{\text{rx},k+\xi_j}]}{2\pi\Delta\nu_C T_s} \quad (14)$$

$$= \begin{cases} \frac{2\pi\Delta\nu_{\text{LS}}T_s + 2\pi\Delta\nu_{\text{LO}}T_s}{2\pi\Delta\nu_C T_s}, & \xi_i = \xi_j \\ \frac{2\pi\Delta\nu_{\text{LS}}T_s}{2\pi\Delta\nu_C T_s}, & \xi_i \neq \xi_j \end{cases} \quad (15)$$

$$= \begin{cases} 1, & \xi_i = \xi_j \\ \Delta\nu_{\text{LS}}/\Delta\nu_C, & \xi_i \neq \xi_j \end{cases} \quad (16)$$

where (12) comes since $\theta_{w,k}^{(i)} - \theta_{w,k-1}^{(i)}$ and $\theta_{w,k}^{(j)} - \theta_{w,k-1}^{(j)}$ are zero mean. Moreover, (13) follows from (2)–(4), (14) and (15) are obtained since the random-walk innovations are independent of each other across time and channels, and (16) comes due to the definition of $\Delta\nu_C$. Finally, (11) is obtained from (16) by noting that $\Delta\nu_{\text{LS}}/\Delta\nu_C = R_{\text{LW}}/(1 + R_{\text{LW}})$.

REFERENCES

- [1] G. Colavolpe, A. Barbieri, and G. Caire, “Algorithms for iterative decoding in the presence of strong phase noise,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.
- [2] D. S. Millar, R. Maher, D. Lavery, T. Koike-Akino, M. Pajovic, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, “Design of a 1 Tb/s superchannel coherent receiver,” *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1453–1463, Mar. 2016.
- [3] T. Pfau, S. Hoffmann, and R. Noé, “Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations,” *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 2009.
- [4] E. Ip and J. M. Kahn, “Feedforward carrier recovery for coherent optical communications,” *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.

- [5] A. F. Alfredsson, E. Agrell, H. Wymeersch, B. J. Puttnam, G. Rademacher, R. S. Luís, and M. Karlsson, “Pilot-aided joint-channel carrier-phase estimation in space-division multiplexed multicore fiber transmission,” *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1133–1142, Feb. 2019.
- [6] L. Lundberg, M. Mazur, A. Lorences-Riesgo, M. Karlsson, and P. A. Andrekson, “Joint carrier recovery for DSP complexity reduction in frequency comb-based superchannel transceivers,” in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. Th.1.D.3.
- [7] M. P. Yankov, L. Barletta, and D. Zibar, “Phase noise compensation for nonlinearity-tolerant digital subcarrier systems with high-order QAM,” *IEEE Photon. J.*, vol. 9, no. 5, Oct. 2017.
- [8] M. D. Feuer, L. E. Nelson, X. Zhou, S. L. Woodward, R. Isaac, B. Zhu, T. F. Taunay, M. Fishteyn, J. M. Fini, and M. F. Yan, “Joint digital signal processing receivers for spatial superchannels,” *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1957–1960, Nov. 2012.
- [9] D. J. Richardson, J. M. Fini, and L. E. Nelson, “Space-division multiplexing in optical fibres,” *Nat. Photon.*, vol. 7, no. 5, pp. 354–362, Apr. 2013.
- [10] B. J. Puttnam, J. Sakaguchi, J.-M. Delgado Mendinueta, W. Klaus, Y. Awaji, N. Wada, A. Kanno, and T. Kawanishi, “Investigating self-homodyne coherent detection in a 19 channel space-division-multiplexed transmission link,” *Opt. Express*, vol. 21, no. 2, pp. 1561–1566, Jan. 2013.
- [11] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, “Influence and suppression of phase noise in multi-antenna OFDM,” in *Proc. Vehicular Technol. Conf.*, vol. 2, Sep. 2004, pp. 1443–1447.
- [12] R. Krishnan, G. Colavolpe, A. Graell i Amat, and T. Eriksson, “Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels,” *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3360–3375, Jul. 2015.
- [13] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, “An EM-based phase-noise estimator for MIMO systems,” in *Proc. IEEE International Conf. Commun.*, Jun. 2013, pp. 3215–3219.
- [14] A. F. Alfredsson, R. Krishnan, and E. Agrell, “Joint-polarization phase-noise estimation and symbol detection for optical coherent receivers,” *J. Lightw. Technol.*, vol. 34, no. 18, pp. 4394–4405, Sep. 2016.
- [15] A. F. Alfredsson, E. Agrell, and H. Wymeersch, “Iterative decoding and phase-noise compensation for multichannel optical transmission,” *IEEE Trans. Commun.*, Apr. 2019, early access.
- [16] R. S. Luís, B. J. Puttnam, J.-M. Delgado Mendinueta, W. Klaus, Y. Awaji, and N. Wada, “Comparing inter-core skew fluctuations in multi-core and single-core fibers,” in *Proc. Conf. Lasers and Electro-Opt.*, May 2015, p. SM2L.5.
- [17] B. J. Puttnam, R. S. Luís, G. Rademacher, A. Alfredsson, W. Klaus, J. Sakaguchi, Y. Awaji, E. Agrell, and N. Wada, “Characteristics of homogeneous multi-core fibers for SDM transmission,” *APL Photon. J.*, vol. 4, no. 2, p. 022804, Feb. 2019.
- [18] R. S. Luís, B. J. Puttnam, J.-M. Delgado Mendinueta, Y. Awaji, and N. Wada, “Impact of spatial channel skew on the performance of spatial-division multiplexed self-homodyne transmission systems,” in *Proc. International Conf. Photon. Switching*, Sep. 2015, pp. 37–39.
- [19] L. Lundberg, B. J. Puttnam, R. S. Luís, G. Rademacher, M. Karlsson, P. A. Andrekson, Y. Awaji, and N. Wada, “Master-slave carrier recovery for M-QAM multicore fiber transmission,” to appear in *Opt. Express*.
- [20] A. Lorences-Riesgo, T. A. Eriksson, A. Fülöp, P. A. Andrekson, and M. Karlsson, “Frequency-comb regeneration for self-homodyne superchannels,” *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1800–1806, Apr. 2016.
- [21] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, “On the impact of intercore skew on joint-core carrier-phase estimation,” in *Proc. European Conf. Opt. Commun.*, Sep. 2019, p. Tu.3.B.4.
- [22] J. C. M. Diniz, F. Da Ros, E. P. da Silva, R. T. Jones, and D. Zibar, “Optimization of DP-M-QAM transmitter using cooperative coevolutionary genetic algorithm,” *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2450–2462, Jun. 2018.
- [23] I. Fatadin, D. Ives, and S. J. Savory, “Blind equalization and carrier phase recovery in a 16-QAM optical coherent system,” *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3042–3049, Aug. 2009.
- [24] S. L. Miller and D. Childers, *Probability and Random Processes: With Applications to Signal Processing and Communications*, 2nd ed. Cambridge, MA, USA: Academic Press, 2012.
- [25] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, “Pilot distributions for phase tracking in space-division multiplexed systems,” in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. P1.SC3.48.